

REPRO_EXTRACT: System Overview

1 Introduction

The *Grosse Reprotox* system is a collection of FORTRAN programs which reside on a VAX/OpenVMS platform. The action of these programs is to create and manage a collection of flat files, some system related and others study related, which together hold and describe information obtained from both GLP and non-GLP reproductive toxicology studies.

Over the years, *Reprotox* data have been instrumental in forming and supporting decisions made by user organisations in safety trials associated with the development of new products. Such organisations have a regulatory obligation to preserve safety data during the data retention period. They also have a business need to maximise their investment in the data for the benefit of future studies.

The data retention period can run into decades, far outliving the normal lifetime of the systems used for the capture, storage and presentation of experimental data. Ultimately, through obsolescence, dwindling skill base and rising support costs, the decision to keep originating systems as the vehicle for maintaining such data becomes untenable. The REPRO_EXTRACT application, which enables data to be analysed in ways not possible using the originating system, extracts data into an 'open' format, allowing the originating system to be retired while maximising the choices for subsequent data analysis and storage.

The REPRO_EXTRACT application is designed to fulfil the primary objective of extracting data from *Grosse Reprotox* systems in a way which can be validated and in so doing give absolute confidence that the originating system can be decommissioned without the loss of critical data.

2 Data Extraction Overview

The REPRO_EXTRACT application is a software tool written specifically to convert *Reprotox* data from internal (VMS binary) format into universal (ANSI) structures that are accessible by any computer system. It comprises two major elements: 1) a user interface for selecting studies; and 2) an extraction engine for performing all data extraction activities. The features of these components are given below:

2.1 User Interface

The user interface has the following features:

REPRO_EXTRACT: System Overview

- Accessible via a special REPRO_EXTRACT OpenVMS account
- Screen (character cell) based (native/emulation)
- Non-scrolling data entry form with online help
- Study selection (individual/bulk)
- Display study title and protocolled activities (individually selected studies only)
- Submit user selections as a parameter list to the extraction engine (batch)

A parameter list, the contents of which is dependent on the selected study's protocolled activities, is submitted to the batch queue associated with the extraction engine and the screen is re-initialised in readiness to receive the next user selection. Exit to the command line (DCL) is supported. The user interface does not provide any information relating to the activity of the extraction engine.

2.2 Extraction Engine

The extraction engine has the following features:

- Runs in batch with extraction instructions supplied as a parameter list
- Creates and preserves the batch log as an audit of program activity
- Creates and preserves an extraction log to record all study extraction events
- Provides an extraction activity cross-check to identify extraction inconsistencies
- Creates and preserves an error log (normal entry = "Extraction successful")
- Transfers automatically all study data (extraction files and pre-existing edit and backup ASCII files) to a Windows platform.

The extraction engine runs from the REPRO_EXTRACT batch queue.

Extraction actions are determined by the contents of the passed parameter list. At the end of the extraction process, all relevant files are transferred to a target directory on a Windows platform. This can be via FTP, PATHWORKS or other suitable automated method (configurable).

3 Reprotox Files Supported

The following tables show the name and location of all files that are accessed by REPRO_EXTRACT (either as part of the extraction process or as part of the subsequent file transfer).

REPRO_EXTRACT: System Overview

3.1 System Files

The following files have system-wide scope. The contents of these files are captured within the extracted study data. As a consequence, there is no added value in extracting them as stand alone entities, particularly as they do not meet the criteria for extraction (i.e. not Study related). The only exception is ZZ.PSW (necessary for identifying users within the edit (<study>.EDI) and backup (<study>.BUP and <study>.LAT) ASCII files):

Filename	Function / Usage	Extract ?
R Module:		
ZR1:ACT.MFI	Gestation/lactation activity reference lookup	x
ZR1:PRO.MFI	Gestation schedule lookup	x
Z Module:		
ZZ1:ACT.MFI	Pre-mating activity reference lookup	x
ZZ1:ZSP.MFI	Species/strain reference lookup	x
ZZ1:ZZ.PSW	User information and operator passwords	✓

3.2 Study Files (Binary)

The following files have study-wide scope and, with the exception of the *Fetal Numeric Protocol, Exchange* and *Cage Mates* files, are the source for data extraction:

Filename	Function / Usage	Extract ?
R Module:		
ZR1:<study>.B	Fetal bodyweight and Gravid Uterine weight data	✓
ZR1:<study>.BP	Male mating data	✓
ZR1:<study>.nnB	Physical development on day nn	✓
ZR1:<study>.CESC	Cauda epididymal sperm count	✓
ZR1:<study>.DVM	Dosing data	✓
ZR1:<study>.ES	Estrous data	✓
ZR1:<study>.FDD	Food added/spilt (note: this is an ASCII file)	✓
ZR1:<study>.G	Female bodyweight data (during gestation)	✓
ZR1:<study>.H	Female bodyweight data (during lactation)	✓
ZR1:<study>.L	C-section data	✓
ZR1:<study>.nnL	Pup bodyweight data on lactation day nn	✓
ZR1:<study>.N	Fetal numeric protocol (max.=5 parameters; 0-4)	x
ZR1:<study>.n (n=0-4)	Fetal data (0=crown-rump length, etc. as per protocol)	✓
ZR1:<study>.ORWT	Organ weights	✓
ZR1:<study>.PGS	Pregnancy status	✓
ZR1:<study>.PPG	Maternal gross findings examination protocol	✓
ZR1:<study>.RPG	Maternal gross findings (pointers/comments) at necropsy	✓
ZR1:<study>.SPG	Maternal gross findings (text) at necropsy	✓

REPRO_EXTRACT: System Overview

Filename	Function / Usage	Extract ?
R Module:		
(continued)		
ZR1:<study>.PPN	Pup gross findings examination protocol	✓
ZR1:<study>.RPN	Pup gross findings (pointers/comments)	✓
ZR1:<study>.SPN	Pup gross findings (text)	✓
ZR1:<study>.PEX	Fetal gross (external) findings examination protocol	✓
ZR1:<study>.REX	Fetal gross (external) findings (pointers/comments)	✓
ZR1:<study>.SEX	Fetal gross (external) findings (text)	✓
ZR1:<study>.PPV	Pup visceral findings examination protocol	✓
ZR1:<study>.RPV	Pup visceral findings (pointers/comments)	✓
ZR1:<study>.SPV	Pup visceral findings (text)	✓
ZR1:<study>.PVI	Fetal visceral findings examination protocol	✓
ZR1:<study>.RVI	Fetal visceral findings (pointers/comments)	✓
ZR1:<study>.SVI	Fetal visceral findings (text)	✓
ZR1:<study>.PSK	Fetal (or pup) skeletal findings examination protocol	✓
ZR1:<study>.RSK	Fetal (or pup) skeletal findings (pointers/comments)	✓
ZR1:<study>.SSK	Fetal (or pup) skeletal findings (text)	✓
ZR1:<study>.nnSM	Sexual maturity on day nn	✓
ZR1:<study>.SMOR	Sperm morphology	✓
ZR1:<study>.SMOT	Sperm motility	✓
ZR1:<study>.TIDC	Sperm head count	✓
ZR1:<study>.VF	Formulation	✓
ZR1:<study>.WM	Water maze	✓
ZR1:<study>.X	Female food fed/left weight data (during gestation)	✓
ZR1:<study>.Y	Female food fed/left weight data (during lactation)	✓
ZR1:<study>.Z	Gestation/lactation activities and gestation schedule	✓
T Module:		
ZT1:<study>.O	Clinical observations (pointers)	✓
ZT1:<study>.L	Clinical observations (text/comments)	✓
ZT1:<study>.nnn	Food & body weights (during pre-mating) on day nnn	✓
ZT1:<study>.X	Exchange file (study migration)	x
Z Module:		
ZZ1:<study>.MTR	Study master protocol	✓
ZZ1:<study>.ANO	Study animals	✓
ZZ1:<study>.CM	Cage mates	x
ZZ1:<study>.FOT	Study footnotes	✓
ZZ1:<study>.LOG	Study log	✓

3.3 Study Files (ASCII)

All study related ASCII files (.EDI, .BAL, .BUP, & .LAT) contain valuable information which needs to be preserved (by copying to Windows for permanent storage):

REPRO_EXTRACT: System Overview

Filename	Function / Usage	Export ?
R Module:		
ZR1:<study>.EDI	Input module edits	✓
ZR1:<study>.BAL	Balance calibration data	✓
ZR1:<study>.FDD	Food added weights	✓
T Module:		
ZT1:<study>.EDI	Input module edits	✓
Z Module:		
ZZ1:<study>.EDI	Input module edits	✓
Backup Data:		
BUP:<study>.BUP	Backup Weight data	✓
BUP:<study>.LAT	Backup Clinical Observations	✓

Note: The above tables are not exhaustive; only those files that are accessed by *Repro-tox* reporting programs are candidates for REPRO_EXTRACT and all other files are ignored.

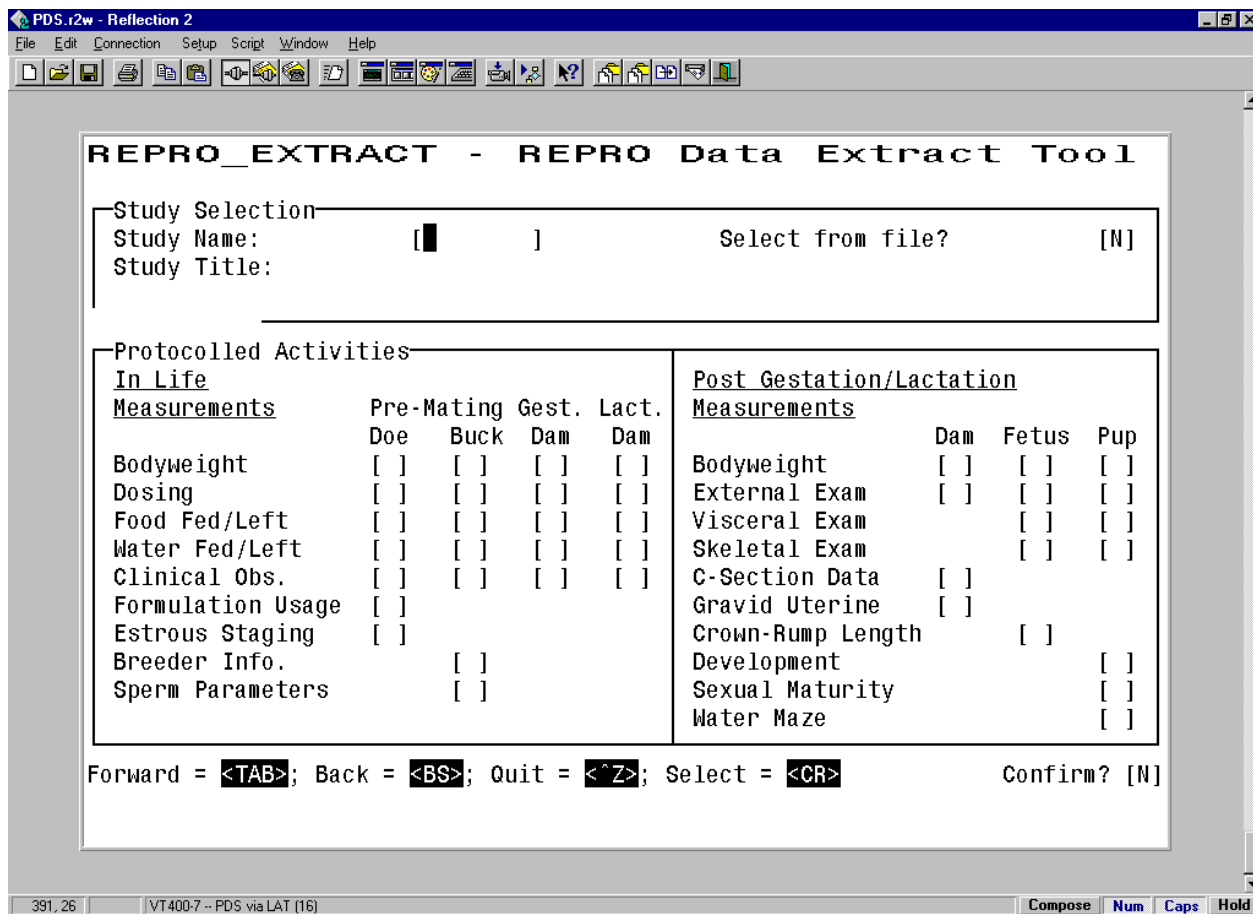
Because of the approach that Scientific Software Systems (the *Repro-tox* system vendor) took in supplying a product that could be tailored to the specific needs of its customers, the REPRO_EXTRACT product is of a highly modular design which lends itself to customisation in order to meet the requirements of each individual installation.

4 User Interface

The purpose of the user interface is to enable *Repro-tox* studies to be selected for extraction. For each individually entered study, key information is displayed prior to confirmation to indicate that the selection exists. Conversely, for studies entered from a file, information is displayed at the time of extraction to indicate that study submission has either been successful or was rejected.

Every time a user logs in to the REPRO_EXTRACT account the *User Interface* is presented:

REPRO_EXTRACT: System Overview



5 Extraction Files

For each iteration, the extraction process creates a number of new (extraction) files on a per study basis which are directed to REPRO_EXTRACT_OUTPUT:[<study>.DATA] on the VAX. All files in this directory contain tabular data (i.e. many rows with each row holding a single datum). Files of supporting information, which provide a record of extraction events, are directed to REPRO_EXTRACT_OUTPUT:[<study>.LOG] on the VAX. All files in the [.DATA] and [.LOG] directories are subsequently copied to Windows.

Backup and Edit files do not require their data to be 'extracted' but they do, however, need to be preserved by moving (copying) from the VAX system. Edit files can exist in each *Reprotox* module (i.e. R = Reproduction; T = Toxicology; and Z = System). These are copied to the EDIT folder (\<study>EDIT) on the target system. In order to preserve their origin, each file is prefixed by their module IDs (e.g. ZRn:<study>.EDI is saved as R_<study>.EDI). Balance Calibration (BAL) and Food Added/Spilt (FDD) files are also copied to the EDIT folder.

REPRO_EXTRACT: System Overview

In a similar way to edit files, backup (BUP & LAT) files are directed to the BACKUP folder (\\<study>\BACKUP) on the target system. In order to preserve file versioning, which isn't supported on the Windows platform, the original version numbers are included in the filename (e.g. <study>.BUP;123 is saved as <study>_123.BUP).

The following tabulation shows all possible files and their target locations on Windows.

Filename	Filename
Extraction files: (\\<study>\DATA)	
<study>_PROTOCOL.CSV	<study>_GROUPS.CSV
<study>_ANIMALS.CSV	<study>_FOOTNOTES.CSV
<study>_OPERATORS.CSV	<study>_ACTIVITIES.CSV
<study>_TEST_PERIODS.CSV	<study>_DOSING.CSV
<study>_FORMULATION_DATA.CSV	<study>_ESTROUS_STAGING.CSV
<study>_<entity>_CLINICAL_OBS.CSV	<study>_<phase>_BODYWEIGHTS.CSV
<study>_<entity>_BODYWEIGHTS.CSV	<study>_<phase>_FOOD_WEIGHTS.CSV
<study>_<entity>_NECROPSY_EXAM.CSV	<study>_<entity>_EXTERNAL_EXAM.CSV
<study>_<entity>_VISCERAL_EXAM.CSV	<study>_<entity>_SKELETAL_EXAM.CSV
<study>_C-SECTION_DATA.CSV	<study>_GRAVID_UTERINE_WEIGHTS.CSV
<study>_FETAL_CROWN-RUMP_LENGTHS.CSV	<study>_SPERM_DATA.CSV
<study>_PUP_STATUS.CSV	<study>_PUP_DEVELOPMENT_DATA.CSV
<study>_PUP_MATURITY_DATA.CSV	<study>_PUP_WATER_MAZE_DATA.CSV
<study>_STUDY_LOG.CSV	REPRO_SYSTEM_OPERATORS.CSV
Log files: (\\<study>\LOG)	
<study>_BATCH.LOG	
<study>_ERROR.LOG	
<study>_EXTRACT.LOG	
Edit files: (\\<study>\EDIT)	
R_<study>.EDI	
T_<study>.EDI	
Z_<study>.EDI	
Miscellaneous files: (\\<study>\EDIT)	
<study>.BAL	
<study>.FDD	
Backup files: (\\<study>\BACKUP)	
<study>_<vvv>.BUP	
<study>_<vvv>.LAT	
Where:	
<study>	= 8 character study name
<phase>	= PRE-MATING, GESTATION, LACTATION
<entity>	= PARENTAL, FETAL, PUP
<vvv>	= version no. (001-999)

REPRO_EXTRACT: System Overview

6 Example Extraction File: C-Section Data

Extraction file(s):	<study>_C-SECTION_DATA.CSV
Source file(s):	ZR1:<study>.L

Field Number	Description	
1	Animal number	}
2	Animal name	
3	Group number	
4	Group name	
5	Calendar day of C-section (days from 1/1/73)	
6	Date of C-section (dd-mmm-yyyy)	
7	Phase of C-section (Premating, Gestation, Lactation)	
8	Phase day of C-section	
9	Operator code	
10	Operator name	
11	Corpora lutea count (left ovary)	
12	Corpora lutea count (right ovary)	
13	Number of nidations (implants)	
14	Cervix position (number of implants to the left)	
15	Number of fetuses	
16	Fetus number	}
17	Gender code of fetus (0 - 3)	
18	Gender text	
19	Status of fetus (1 - 7)	
20	Status text	
21	Fixative code of fetus (0 - 3)	
22	Fixative text	

The following example is a fragment of a C-Section extract file (where the heading row is word-wrapped for clarity). Note the interleaving of dam and fetus records:

```
Animal #|Animal name|Group #|Group name|Entry day|Entry date|Phase|Phase day|Op code|Op
name|Corpora lutea left|Corpora lutea right|# nidations|Cervix position|# fetuses|Fetus
#|Gender|Gender text|Status|Status text|Fixative code|Fixative text
1|101|1|CONTROL|11972|11-OCT-2005|Gestation|21|9|<OPERATOR>|6|7|13|6|13| || || |
1|101|1|CONTROL|11972|11-OCT-2005|Gestation|21|9|<OPERATOR>| || || |1|2|Female|1|ALIVE|2|VISCERAL
1|101|1|CONTROL|11972|11-OCT-2005|Gestation|21|9|<OPERATOR>| || || |2|2|Female|1|ALIVE|1|SKELETAL
1|101|1|CONTROL|11972|11-OCT-2005|Gestation|21|9|<OPERATOR>| || || |3|2|Female|1|ALIVE|2|VISCERAL
1|101|1|CONTROL|11972|11-OCT-2005|Gestation|21|9|<OPERATOR>| || || |4|2|Female|1|ALIVE|1|SKELETAL
1|101|1|CONTROL|11972|11-OCT-2005|Gestation|21|9|<OPERATOR>| || || |5|2|Female|1|ALIVE|2|VISCERAL
1|101|1|CONTROL|11972|11-OCT-2005|Gestation|21|9|<OPERATOR>| || || |6|2|Female|1|ALIVE|1|SKELETAL
1|101|1|CONTROL|11972|11-OCT-2005|Gestation|21|9|<OPERATOR>| || || |7|1|Male|1|ALIVE|2|VISCERAL
1|101|1|CONTROL|11972|11-OCT-2005|Gestation|21|9|<OPERATOR>| || || |8|1|Male|1|ALIVE|1|SKELETAL
1|101|1|CONTROL|11972|11-OCT-2005|Gestation|21|9|<OPERATOR>| || || |9|1|Male|1|ALIVE|2|VISCERAL
1|101|1|CONTROL|11972|11-OCT-2005|Gestation|21|9|<OPERATOR>| || || |10|2|Female|1|ALIVE|1|SKELETAL
1|101|1|CONTROL|11972|11-OCT-2005|Gestation|21|9|<OPERATOR>| || || |11|2|Female|1|ALIVE|2|VISCERAL
1|101|1|CONTROL|11972|11-OCT-2005|Gestation|21|9|<OPERATOR>| || || |12|1|Male|1|ALIVE|1|SKELETAL
1|101|1|CONTROL|11972|11-OCT-2005|Gestation|21|9|<OPERATOR>| || || |13|1|Male|1|ALIVE|2|VISCERAL
2|102|1|CONTROL|11972|11-OCT-2005|Gestation|21|9|<OPERATOR>|7|6|13|7|13| || || |
```

etc...